SPECIFICATION

ALIGNMENT SYSTEM AND ALIGNING METHOD FOR MULTILINGUAL DOCUMENTS

FIELD OF THE INVENTION

本発明は、複数の言語で構成される文書間の文書対応付けシステムにかかり、特に、２言語以上で記述された対訳文書の、文の対応付けを行う複数言語文書の対応付けシステム、複数言語文書の対応付け方法、この方法を行わせるプログラム、及びこのプログラムを格納した記録媒体に関する。

BACKGROUND OF THE INVENTION

複数国に輸出されることが予想される製品のマニュアルなどのように、複数の言語で同じ内容の文書を記述する場合が増えている。このような複数の言語文書の対訳の正確性を評価，担保するため、これらの文の対応付けを行う需要も増えている。従来技術文献である“宇津呂 武仁，松本 裕治　共著「対訳辞書及び統計情報を用いた二言語対訳テキスト照合」(「コンピュータソフトウェア」岩波書店　vol.12 No.5 Sep.1995 p.12(414)-p.21(423))”には、対訳文書の文の対応付けを、対訳辞書を利用したダイナミックプログラミングで行う方法が記載されている。

上記従来技術文献によれば、対応付けを行うには、文書を１文毎に区切り、さらにその文の形態素解析を行って、単語毎に分割する。そして、これらの単

語の中から自立語を取り出し、対訳辞書を用いてそれぞれの文の中の自立語がどの程度対応しているか（どの程度意味内容が一致しているか）によって対応付けを評価する。この評価に際して、例えば以下のような式を用いる。

$h(x, y)=2 \times f_m(x, y) / (f_j(x)+f_j(y))$

ここで

$h(x, y)$は評価関数,

x は、原文中の文（複数文の場合もある）,

y は、訳文中の文（複数文の場合もある）,

$f_m(x, y)$は、文 x と文 y の中で対応の付いた自立語の数,

$f_j(x)$は、文 x 中の自立語の数,

$f_j(y)$は、文 y 中の自立語の数,

である。

このような式を用いた評価を行えば、文書の対応の割合が大きいほど評価関数 $h(x, y)$の値は大きくなり（最大値：１）、逆に割合が小さいほど値は小さくなる（最小値：０）。この評価関数を文の先頭から調べていき、評価関数の和が最も大きくなる組合せを、対応付け問題の解とする。

しかしながら、上記の方法では、通常の２言語の対訳文書間の文の対応付けを、３言語以上の文書間の文の対応付けに適用する場合に、

・複数の辞書を利用するため、システムにかなりの量の記録領域を必要とする。

・評価の処理に時間がかかる。

・全ての言語間で、各言語対の対応の整合性をとるのが困難である。

といった問題がある。

また、２言語の対訳文書の対応付けに関しても、高精度での対応を自動的に付けるのは難しく、対応付けの結果を見ながらの人手によるチェックや修正が必要であり、その作業工数の発生が問題となっている。

本発明は、従来の複数言語文書の対応付けシステムが有する上記問題点に鑑みてなされたものである。そして、本発明の目的は、英語－日本語－ドイツ語など、複数の言語でそれぞれ構成される文書間の文の対応付けを効率良く行うための、新規かつ改良された複数言語文書の対応付けシステム、及び複数言語文書の対応付け方法を提供することにある。

SUMMARYOF THE INVENTION

上記課題を解決し、複数の言語で構成された、同一内容の文書間の、文の対応付けを効率良く行う複数言語間文書の対応付けシステムを実現するために、本発明の複数言語文書の対応付けシステムは、ｎ種類（：ｎは２以上の自然数）の言語の文書を単語毎に分割する形態素解析手段と、ｎ種類の言語の文書のうちの２種類を選択する手段と、選択された２種類の言語の文書の評価関数を計算する手段と、評価結果に応じてｎ種類の言語の文書を対応付ける手段とを備える。

BRIEF DESCRIPTION OF THE DRAWINGS

図１は、第１の実施例にかかる複数言語文書の対応付けシステムの構成を示す説明図である。

図２は、図１の複数言語文書の対応付けシステムの動作を示すフローチャー

トである。

図３は、第２の実施例にかかる複数言語文書の対応付けシステムの構成を示す説明図である。

図４は、図３の複数言語文書の対応付けシステムの動作を示すフローチャートである。

図５は、第３の実施例にかかる複数言語文書の対応付けシステムの構成を示す説明図である。

図６は、図５の複数言語文書の対応付けシステムの動作を示すフローチャートである。

図７は、第４の実施例にかかる複数言語文書の対応付けシステムの構成を示す説明図である。

図８は、図７の複数言語文書の対応付けシステムの動作を示すフローチャートである。

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS OF THE INVENTION

以下に添付図面を参照しながら、本発明にかかる複数言語文書の対応付けシステムと、このシステムを用いた複数言語文書の対応付け方法に関する、好適な実施例について、以下に詳細に説明する。

（第１の実施例）

図１は、第１の実施例にかかる複数言語文書の対応付けシステム 100 の構成

を示す説明図である。複数言語文書の対応付けシステム 100 は、図 1 に示されるように、文分割手段 105，形態素解析手段 106，評価関数計算手段 107，計算結果管理手段 108，及び対訳辞書データベース 109 から構成されている。この実施例では、各言語のファイル 101～104 が入力されて、対応タグ付きファイル 110～113 が出力される。

以下、各構成要素について詳細に説明する。

英語ファイル 101 は英語で記述された文書ファイル，日本語ファイル 102 は日本語で記述された文書ファイル，ドイツ語ファイル 103 はドイツ語で記述された文書ファイル，そして中国語ファイル 104 は中国語で記述された文書ファイルである。これら 4 つの文書ファイルには、用いられる言語は異なるが同一の内容が記載されて、それぞれが対訳形式になっている。

文分割手段 105 は、文書ファイルを 1 文毎に分割する。例えば、英文であればピリオッド「.」、日本語ならば句点「 」など目安として、文書を 1 文単位に分割する。形態素解析手段 106 は、形態素解析処理を行い、文を単語毎に分割する。文分割手段 105 及び形態素解析手段 106 は、既存の構成を適用可能であり、処理動作の詳細については説明を省略する。

評価関数計算手段 107 は、最適な対応付けを見つけるために、与えられた評価関数を計算する。例えば、評価関数は、次の式；

$$h(x, y)=2 \times f_m(x, y)/(f_j(x)+f_j(y))$$

で表される。ここで、h(x, y)は評価関数，x は一方の言語の文（原文），y は他方の言語の文（訳文），$f_m(x, y)$は文 x と文 y の中で対応の付いた自立語の数，$f_j(x)$は文 x 中の自立語の数，$f_j(y)$は文 y 中の自立語の数である。

計算結果管理手段108は、評価関数計算手段が計算した結果を保持し、既出の評価関数計算が再び到来したときに保持している結果を出力し、同じ計算を何度も行わないようにする。

対訳辞書データベース109は、対応付けをするための辞書で、原文の単語を引くと、訳文の語が1つまたは複数あるような辞書である。例えば、原文が英語、訳文が日本語の場合、英和辞書に相当する。

対応タグ付き英語ファイル110は、英語ファイル101に、文書中の各文が他の文書のどの文に対応しているのかを示すタグを付与したものである。対応タグ付き日本語ファイル111，対応タグ付きドイツ語ファイル112，及び対応タグ付き中国語ファイル113も同様に、各々、元の日本語ファイル102，ドイツ語ファイル103および中国語ファイル104に、文書中の各文が他の文書のどの文に対応しているのかを示すタグを付与したものである。

本実施例にかかる複数言語文書の対応付けシステム100は、上記の通り構成されている。次に、図2を参照しながら、複数言語文書の対応付けシステム100の動作を説明する。

図2は、この複数言語文書の対応付けシステム100の動作を示すフローチャートである。

ステップS10では、文分割手段105によって、一方（原文）の文書ファイルと他方（訳文）の文書ファイルの各々について文分割を行う。そして、対応付けをどこまで行ったかを示すカウンタNを、0にセットする。

ステップS11では、カウンタNをインクリメント（＋1）する。

ステップS12では、対応付けを行う言語の数がカウンタNと等しいかどうか

を比較する。もしも等しければ、ステップS17に行く。

ステップS13では、対応付けを行う言語をN番目とN+1番目にセットする。

ステップS14では、評価関数計算手段105がセットされた言語に対して文の対応付けを行う。

ステップS15では、対応付けを行った結果に対して、対応する文同士に双方向リンクを張る。

ステップS16では、２対１，３対１などの複数文の対応になってしまった分に対してマーク付けを行う。これらのマーク付けされた文の組は、次に対応付けを行う場合は、それを主文とみなして処理する。

ステップS17では、対応付けを行っていない言語同士の文に対して、他の言語同士の対応付け結果を利用して、リンクを張る。

以上の処理を、図１の４言語（ｎ＝４）間の対応付けを行う場合を例にとって説明する。この例では、英語が１番目，日本語が２番目，ドイツ語が３番目，そして中国語が４番目の言語に相当する。

まず、４つの言語それぞれを文分割手段105によって一文毎に分割する。

次に、文の対応付けを行う。英語と日本語の対応付けは英日対訳辞書を使って，日本語とドイツ語の対応付けは日独対訳辞書を使って、ドイツ語と中国語の対応付けは対訳辞書を使ってそれぞれ行う。これにより、英語－日本語間，日本語－ドイツ語間，ドイツ語－中国語間の合計（ｎ－１）通りの文同士のリンクが生成される。

さらに、対応のついていない言語同士（ここでは、日本語－中国語，英語－ドイツ語，英語－中国語）の文のリンクを張ることによって、すべての言語間

の対応を取ることができる。

以上説明したように、本実施例によれば、対応付けの精度は多少落ちるが、少ない記憶容量で時間もあまりかからずに、効率良く文の対応を。とることができる。

（第２の実施例）

図３に、第２の実施例の複数言語文書の対応付けシステム200の構成を示す。

英語ファイル201は英語で記述された文書ファイル，日本語ファイル202は日本語で記述された文書ファイル，ドイツ語ファイル 203 はドイツ語で記述された文書ファイル，そして中国語ファイル 204 は中国語で記述された文書ファイルである。これら４つの文書ファイルには、用いられる言語は異なるが同一の内容が記載され、それぞれが対訳形式になっている。

文分割手段205は、文書ファイルを１文毎に分割する。例えば、英文であればピリオッド「.」、日本語ならば句点「 」など目安として、文書を１文単位に分割する。形態素解析手段 206 は、形態素解析処理を行い、文を単語毎に分割する。文分割手段 205 及び形態素解析手段 206 は、既存の構成を適用可能であり、処理動作の詳細については説明を省略する。

評価関数計算手段207は、最適な対応付けを見つけるために、与えられた評価関数を計算する。この評価関数としては、例えば第１の実施例で用いた評価関数の式が適用できる。

計算結果管理手段 208 は、評価関数計算手段 207 が計算した結果を保持し、既出の評価関数計算が再び到来したときに保持している結果を出力し、同じ計

算を何度も行わないようにする。

対訳辞書データベース209は、対応付けをするための辞書で、原文の単語を引くと、訳文の語が1つまたは複数あるような辞書である。例えば、原文が英語、訳文が日本語の場合、英和辞書に相当する。

対応タグ付き英語ファイル210は、英語ファイル201に、文書中の各文が他の文書のどの文に対応しているのかを示すタグを付与したものである。対応タグ付き日本語ファイル211，対応タグ付きドイツ語ファイル212，及び対応タグ付き中国語ファイル213も同様に、各々、元の日本語ファイル202，ドイツ語ファイル203および中国語ファイル204に、文書中の各文が他の文書のどの文に対応しているのかを示すタグを付与したものである。

相違箇所表示手段220は、対応付け結果に不整合があった場合に、その不整合箇所を表示し、ユーザに修正させる機能をもつ。不整合とは、例えば、英語の文Enと日本語の文Jnが対応していて、上記日本語の文Jnとドイツ語の文Dnが対応しているときに、英語とドイツ語の対応結果をみると、上記英語の文Enと上記ドイツ語の文Dnとが対応していないような場合である。

図4は、本実施例の複数言語文書の対応付けシステム200の動作を示すフローチャートである。

ステップS20では、文分割手段205によって、一方（原文）の文書ファイルと他方（訳文）の文書ファイルの各々について文分割を行う。そして、対応付けをどこまで行ったかを示すカウンタNとMを、1にセットする。

ステップS21では、対応付けを行う言語の数がカウンタNと等しいかどうかを比較する。もしも等しければ、ステップS27に行く。

ステップS22では、カウンタMをインクリメントし、カウンタNの値をM+1にする。

ステップS23では、対応付けを行う言語の数がカウンタMと等しいかどうかを比較する。もしも等しければ、ステップS28に行く。

ステップS24では、対応付けを行う言語をM番目とN番目にセットする。

ステップS25では、評価関数計算手段205がセットされた言語に対して文の対応付けを行う。

ステップS26では、対応付けを行った結果に対して、対応する文同士に双方向リンクを張る。

ステップS27では、Nをインクリメントする。

ステップS28では、文の対応に不整合がある部分を表示し、ユーザに修正させる。

ステップS29では、ユーザの修正に応じて、対応付けのリンクを張り直す。

このようにして、n種類の言語の文に対して、全ての組合せ（この実施例では、言語の種類n＝4でn(n－1)／2＝6通り)の対応付けを行う。

以上説明したように、本実施例によれば、ユーザの修正処理が発生することが必須となるが、高精度の対応付けが効率良く実現可能となる。

（第3の実施例）

図5に、第3の実施例の複数言語文書の対応付けシステム300の構成を示す。

英語ファイル301は英語で記述された文書ファイル，日本語ファイル302は日本語で記述された文書ファイル，ドイツ語ファイル303はドイツ語で記述さ

れた文書ファイル，そして中国語ファイル 304 は中国語で記述された文書ファイルである。これら４つの文書ファイルには、用いられる言語は異なるが同一の内容が記載され、それぞれが対訳形式になっている。

文分割手段 305 は、文書ファイルを１文毎に分割する。例えば、英文であればピリオッド「.」、日本語ならば句点「 」など目安として、文書を１文単位に分割する。形態素解析手段 306 は、形態素解析処理を行い、文を単語毎に分割する。文分割手段 305 及び形態素解析手段 306 は、既存の構成を適用可能であり、処理動作の詳細については説明を省略する。

評価関数計算手段 307 は、最適な対応付けを見つけるために、与えられた評価関数を計算する。この評価関数としては、例えば第１の実施例で用いた評価関数の式が適用できる。

計算結果管理手段 308 は、評価関数計算手段 307 が計算した結果を保持し、既出の評価関数計算が再び到来したときに保持している結果を出力し、同じ計算を何度も行わないようにする。

対訳辞書データベース 309 は、対応付けをするための辞書で、原文の単語を引くと、訳文の語が１つまたは複数あるような辞書である。例えば、原文が英語、訳文が日本語の場合、英和辞書に相当する。

対応タグ付き英語ファイル 310 は、英語ファイル 301 に、文書中の各文が他の文書のどの文に対応しているのかを示すタグを付与したものである。対応タグ付き日本語ファイル 311，対応タグ付きドイツ語ファイル 312，及び対応タグ付き中国語ファイル 313 も同様に、各々、元の日本語ファイル 302，ドイツ語ファイル 303 および中国語ファイル 304 に、文書中の各文が他の文書のどの文に

対応しているのかを示すタグを付与したものである。

図６は、本実施例の複数言語文書の対応付けシステム 300 の動作を示すフローチャートである。

ステップ S30 では、文分割手段 305 によって、一方（原文）の文書ファイルと他方（訳文）の文書ファイルの各々について文分割を行う。そして、対応付けをどこまで行ったかを示すカウンタ N と M を、１にセットする。

ステップ S31 では、対応付けを行う言語の数がカウンタ N と等しいかどうかを比較する。もしも等しければ、ステップ S37 に行く。

ステップ S32 では、カウンタ M をインクリメントし、カウンタ N の値を M+1 にする。

ステップ S33 では、対応付けを行う言語の数がカウンタ M と等しいかどうかを比較する。もしも等しければ、ステップ S37 に行く。

ステップ S34 では、対応付けを行う言語を M 番目と N 番目にセットする。

ステップ S35 では、評価関数計算手段 305 がセットされた言語に対して文の対応付けを行う。

ステップ S36 では、N をインクリメントする。

ステップ S37 では、対応付けのポイントの和が最も大きくなるような文の組を選択する。

ステップ S38 では、対応する文同士に双方向リンクを張る。

以上の処理を、図５の４言語（ｎ＝４）間の対応付けを行う場合を例にとって説明する。この例では、英語が１番目，日本語が２番目，ドイツ語が３番目，そして中国語が４番目の言語に相当する。

まず、４つの言語それぞれを文分割手段305によって一文毎に分割する。次に、すべての文書の組の評価関数を計算する。この場合、英語－日本語間，英語－ドイツ語間，英語－中国語間，日本語－ドイツ語間，日本語－中国語間，及びドイツ語－中国語間の６つの評価関数を計算する。

次に、対応付けポイントの和が最も大きくなるように対応をとっていく。この対応は、４言語まとめて同時に行われる。例えば、英文１文，日本文１文，ドイツ文２文，中国文１分の評価ポイントは、英文と日本文の１文対１文，英文とドイツ文の１文体２文，英文と中国文の１文対１文，日本文とドイツ文の１文対２文，日本文と中国文の１文対１文，ドイツ文と中国文の２文対１文の評価ポイントの和となる。この計算を続け、評価ポイントの輪が最も大きくなったものを対応付けの正解とする。

以上説明したように、本実施例によれば、処理時間は増加するが、高精度の対応付けが効率良く実現できる。

（第４の実施例）

図７に、第４の実施例の複数言語文書の対応付けシステム400の構成を示す。

英語ファイル401は英語で記述された文書ファイル，日本語ファイル402は日本語で記述された文書ファイル，ドイツ語ファイル403はドイツ語で記述された文書ファイル，そして中国語ファイル404は中国語で記述された文書ファイルである。これら４つの文書ファイルには、用いられる言語は異なるが同一の内容が記載され、それぞれが対訳形式になっている。

文分割手段405は、文書ファイルを１文毎に分割する。例えば、英文であれ

ばピリオッド「.」、日本語ならば句点「 」など目安として、文書を1文単位に分割する。形態素解析手段 406 は、形態素解析処理を行い、文を単語毎に分割する。文分割手段 405 及び形態素解析手段 406 は、既存の構成を適用可能であり、処理動作の詳細については説明を省略する。

評価関数計算手段 407 は、最適な対応付けを見つけるために、与えられた評価関数を計算する。この評価関数としては、例えば第1の実施例で用いた評価関数の式が適用できる

計算結果管理手段 408 は、評価関数計算手段 407 が計算した結果を保持し、既出の評価関数計算が再び到来したときに保持している結果を出力し、同じ計算を何度も行わないようにする。

対訳辞書データベース 409 は、対応付けをするための辞書で、原文の単語を引くと、訳文の語が1つまたは複数あるような辞書である。例えば、原文が英語、訳文が日本語の場合、英和辞書に相当する。

対応タグ付き英語ファイル 410 は、英語ファイル 401 に、文書中の各文が他の文書のどの文に対応しているのかを示すタグを付与したものである。対応タグ付き日本語ファイル 411，対応タグ付きドイツ語ファイル 412，及び対応タグ付き中国語ファイル 413 も同様に、各々、元の日本語ファイル 402，ドイツ語ファイル 403 および中国語ファイル 404 に、文書中の各文が他の文書のどの文に対応しているのかを示すタグを付与したものである。

言語類似度データ 420 は、言語同士の文法などがどれだけ似ているかを数値化したものである。言語同士の文法の類似度が高いほど文の対応付けの程度も向上する。そこで、この言語類似度データ 420 には、それぞれの言語対の類似

度の値が、例えば表形式などで記録されている。

図８は、本実施例の複数言語文書の対応付けシステム 400 の動作を示すフローチャートである。

ステップ S40 では、文分割手段 405 によって、一方（原文）の文書ファイルと他方（訳文）の文書ファイルの各々について文分割を行う。そして、対応付けをどこまで行ったかを示すカウンタ N を、０にセットする。

ステップ S41 では、カウンタ N をインクリメントする。

ステップ S42 では、対応付けを行う言語の数がカウンタ N と等しいかどうかを比較する。もしも等しければ、処理を終了する。

ステップ S43 では、言語類似度データ 420 に基づいて、まだ選択されていない言語対中で言語類似度が最も高いものを選択し、選択済のマークを付与する。

ステップ S44 では、言語対に文対応のリンクが張られているかどうか調べる。もしもリンクが張られていれば、ステップ S43 に行く。

ステップ S45 では、評価関数計算手段 407 が、選択された言語に対して文の対応付けを行う。

ステップ S46 では、対応付けを行った結果に対して、対応する文同士に双方向リンクを張る。

ステップ S47 では、２対１，３対１などの複数文の対応になった文に対してマーク付けを行う。これらのマーク付けされた文の組は、次に対応付けを行う場合はそれを１文とみなして処理する。

ステップ S48 では、間接的に対応の付いた言語に対してリンクを張る。例えば、英語－日本語間，英語－ドイツ語間の対応がとれたとすると、これら２つ

の対応を利用して、日本語－ドイツ語間の対応を求めることが可能であり、この求められた日本語－ドイツ語間にも文対応のリンクを張る。

以上説明したように、本実施例によれば、言語類似度データを用意することで、高速で精度の高い対応付けが効率良く実現できる。

上記の４つの実施例の“速度”，“精度”，“使用する記憶容量”を比較すると、表１のようになる。表１においては、「◎」は優良，「○」は良好，「△」は普通を表す。

【表１】

実施例	速度	精度	記録容量	その他
1	◎	△	◎	
2	○	◎	△	ユーザの修正が必要
3	△	◎	△	
4	◎	○	○	言語類似度データが必要

以上、添付図面を参照しながら本発明にかかる複数言語文書の対応付けシステム及び複数言語文書の対応付け方法の好適な実施例について説明したが、本発明はこれらの実施例の構成に限定されるものではない。当業者であれば、特許請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、それらについても当然に本発明の技術的範囲に属する。

例えば、上記第１～４の実施例では、英語，日本語，ドイツ語，及び中国語間の対応付けを示したが、対訳辞書を変えることによって、どのような言語同士の対応もとることができる。

また、上記各実施例では言語数が４言語（ｎ＝４）の例を示したが、２言語以上であれば何言語の対応付けであっても適用可能である。また、第２，第３の実施例では言語数が増えてくると処理時間が非常に遅くなるおそれがあるが、計算する対応組の数を減らすことによって、処理時間の短縮を図ることができる。

なお、本発明の複数言語文書の対応方法は、ソフトウェアプログラムとして記述することもでき、このプログラムは記録媒体に記録することもできる。

以上説明したように、本発明によれば、複数の言語で構成される文書間の文の対応付けを効率良く行う複数言語文書の対応付けシステムを提供できる。

What is claimed is:

1. ｎ種類（：ｎは２以上の自然数）の言語の文書を対応付けるシステムであって、

各言語の文書を単語毎に分割する形態素解析手段と、

前記ｎ種類の言語の文書のうちの２種類を選択する手段と、

前記選択された２種類の言語の文書の評価関数を計算する手段と、

前記２種類の言語の文書の評価結果によって、前記ｎ種類の言語の文書を対応付ける手段と

を含むことを特徴とする複数言語文書の対応付けシステム。

2. 前記形態素解析手段が、

各言語の文書を文毎に分割する手段と、分割された各文をさらに単語毎に分割する手段とからなることを特徴とする、請求項１記載の複数言語文書の対応付けシステム。

3. 前記ｎ種類の言語の文書のうちの２種類を選択する手段が、

前記ｎ種類の言語の文書を任意の順序で並べたときにｋ番目とｋ＋１番目（：ｋは、１からｎ－１までの自然数）の、ｎ－１通りの組合せを選択することを特徴とする、請求項１記載の複数言語文書の対応付けシステム。

4. 前記ｎ種類の言語の文書のうちの２種類を選択する手段が、

ｎ（ｎ－１）／２通りの組合せを選択することを特徴とする、請求項１記載の複数言語文書の対応付けシステム。

5. 前記評価関数で計算した結果を保持する計算結果保持手段を含むことを特徴とする、請求項１記載の複数言語文書の対応付けシステム。

6. 前記評価関数が、次式；

$h(x, y)=2 \times f_m(x, y) / (f_j(x)+f_j(y))$

で表されることを特徴とする請求項１記載の複数言語文書の対応付けシステム。但し、$h(x, y)$は評価関数，x は一方の言語の文（原文），y は他方の言語の文（訳文），$f_m(x, y)$は文 x と文 y の中で対応の付いた自立語の数，$f_j(x)$は文 x 中の自立語の数，$f_j(y)$は文 y 中の自立語の数である。

7. 前記ｎ種類の言語の文書のうちのいずれか３種類以上の言語の文書の対応付けに不整合が生じたときに、前記不整合の箇所を表示する手段を含むことを特徴とする、請求項１記載の複数言語文書の対応付けシステム。

8. 前記評価関数を計算する手段は、前記評価関数の和が最大になるように最適化しながら対応付けを行うことを特徴とする、請求項１記載の複数言語文書の対応付けシステム。

9. 言語間の類似度データを調べながら、対応付けの正解率の高い言語対を指示する手段を含むことを特徴とする、請求項１記載の複数言語文書の対応付けシステム。

10. ｎ種類（：ｎは２以上の自然数）の言語の文書を対応付ける方法であって、

各言語の文書を単語毎に分割する形態素解析ステップと、

前記ｎ種類の言語の文書のうちの２種類を選択するステップと、

前記選択された２種類の言語の文書の評価関数を計算するステップと、

前記２種類の言語の文書の評価結果によって、前記ｎ種類の言語の文書を対応付けるステップと

を含むことを特徴とする複数言語文書の対応付け方法。

11. コンピュータに、請求項１０記載の複数言語文書の対応付け方法を行わせるステップを記述したことを特徴とするプログラム。

ABSTRACT

複数の言語で構成された、同一内容の文書間の、文の対応付けを効率良く行う複数言語間文書の対応付けシステムを実現するために、本発明の複数言語文書の対応付けシステムは、ｎ種類（：ｎは２以上の自然数）の言語の文書を単語毎に分割する形態素解析手段と、ｎ種類の言語の文書のうちの２種類を選択する手段と、選択された２種類の言語の文書の評価関数を計算する手段と、評価結果に応じてｎ種類の言語の文書を対応付ける手段とを備える。